

Analisis Klasifikasi Pelanggan Listrik Rumah Tangga Bersubsidi Kota Surabaya Menggunakan *Support Vector Machine* dan *Naive Bayes Classifier*

Wiwin Yuli Widiawati dan Raden Mohamad Atok

Departemen Statistika, Fakultas Matematika Komputasi dan Sains Data,

Institut Teknologi Sepuluh Nopember (ITS)

e-mail: moh_atok@statistika.its.ac.id

Abstrak—Pemerintah menyediakan dana bagi kelompok masyarakat tidak mampu yang diberikan dalam bentuk subsidi terhadap tarif tenaga listrik untuk pelanggan rumah tangga dengan daya 450 VA dan 900 VA. Berdasarkan data dari Badan Pusat Statistik (BPS) yang telah diolah oleh Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K), dikeluarkan daftar 40% masyarakat Indonesia yang berada dalam kondisi ekonomi terendah. Pada penelitian ini bertujuan untuk melakukan klasifikasi sesuai karakteristik yang dimiliki oleh pelanggan subsidi listrik rumah tangga daya 450 VA dan 900 VA menggunakan metode *Support Vector Machine* dan *Naive Bayes Classifier*. Hasil Perbandingan antara metode SVM dan NBC didapatkan bahwa SVM memberikan hasil yang lebih baik dengan parameter optimal yang diperoleh pada kernel RBF adalah $C = 10$ dan $\gamma = 1$. Kategori 450 VA sebanyak 91,6% pelanggan rumah tangga diklasifikasikan dengan benar dan sisanya 8,4% pelanggan rumah tangga diprediksi masuk ke dalam kategori subsidi 900 VA. Pada kategori subsidi 900 VA sebanyak 81,9% pelanggan rumah tangga diklasifikasikan dengan benar dan sisanya 18,1% pelanggan rumah tangga diprediksi masuk kategori 450 VA.

Kata Kunci—*Naive Bayes Classifier*, *Radial Basis Function*, *Subsidi Listrik*, *Support Vector Machine*.

I. PENDAHULUAN

LISTRIK adalah energi vital bagi keberlangsungan aktivitas manusia baik individu, kelompok masyarakat maupun dunia industri. Energi listrik dapat dimanfaatkan untuk melakukan aktivitas dengan manfaat yang sangat besar dimana berbagai peralatan memenuhi kebutuhan hidup dioperasikan menggunakan energi listrik [1]. Kegiatan masyarakat yang cenderung meningkat dari waktu ke waktu, mendorong peningkatan pengoperasian peralatan dengan tenaga listrik. PT. Perusahaan Listrik Negara (Persero) yang selanjutnya disebut sebagai PLN merupakan perusahaan penyedia energi listrik dengan mengelompokkan pelanggan menjadi 5 golongan berdasarkan keperluan pemakaian [2]. Golongan tersebut antara lain sektor rumah tangga, industri, bisnis, sosial, dan penerangan umum.

Golongan rumah tangga merupakan kelompok pemakai energi listrik paling besar yaitu mencapai 90% dari total pelanggan [3]. Jumlah rumah tangga yang ada di Provinsi Jawa Timur mengalami kenaikan setiap tahunnya, sehingga permintaan terhadap sambungan listrik disektor tersebut juga mengalami peningkatan. Hal ini juga tidak terlepas karena Jawa Timur memiliki jumlah penduduk terpadat kedua di Indonesia [4]. PT PLN (Persero) memiliki wilayah kerja di

seluruh Indonesia, salah satunya adalah PT PLN (Persero) Distribusi Jawa Timur yang menargetkan rasio elektrifikasi di provinsi Jawa Timur dapat mencapai 100% pada tahun 2019 dari posisi saat ini 91%.

Pemerintah mencanangkan Program Indonesia Terang dalam bentuk pembangunan infrastruktur penyediaan tenaga listrik. Pemerintah menyediakan dana bagi kelompok masyarakat miskin dan tidak mampu yang diberikan dalam bentuk subsidi terhadap tarif tenaga listrik untuk pelanggan rumah tangga dengan daya listrik 450 VA dan 900 VA. Besaran subsidi listrik yang diterima konsumen rumah tangga daya 450 VA dan 900 VA tergantung pada jumlah pemakaian energi listriknya. Secara rata-rata, konsumen rumah tangga daya 450 VA mendapatkan subsidi listrik sebesar Rp 80.000 per bulan, dan untuk konsumen rumah tangga daya 900 VA adalah rata-rata Rp 90.000 per bulan [3]. Pembangunan subsidi listrik tersebut membutuhkan anggaran yang sangat besar. Pemerintah perlu memastikan bahwa penerima manfaat subsidi listrik adalah hanya masyarakat miskin dan tidak mampu. Berdasarkan data dari Badan Pusat Statistik (BPS) yang telah diolah oleh Tim Nasional Percepatan Penanggulangan Kemiskinan (TNP2K), dikeluarkan daftar 40% masyarakat Indonesia yang berada dalam kondisi ekonomi terendah [5]. Data yang dikeluarkan oleh TNP2K tersebut menjadi database rumah tangga yang menjadi prioritas program penanggulangan kemiskinan di Indonesia. Pada penelitian ini bertujuan untuk melakukan klasifikasi sesuai karakteristik yang dimiliki oleh pelanggan subsidi listrik rumah tangga daya 450 VA dan 900 VA dengan menerapkan metode data mining. PLN merupakan salah satu perusahaan yang bisa memanfaatkan metode data mining karena permintaan sambungan listrik oleh pelanggan yang berlangsung setiap hari sehingga semakin lama data pelanggan yang tersimpan sangatlah besar.

Data mining merupakan suatu proses untuk identifikasi pola yang memiliki potensi dan berguna untuk mengelola dataset yang besar [6]. Dalam data mining terdapat 10 algoritma teratas yang paling berpengaruh dan dipilih oleh peneliti dalam komunitas data mining, dimana lima diantaranya adalah algoritma klasifikasi yaitu *Support Vector Machines* (SVM), *AdaBoost*, *K-Nearest Neighbor* (K-NN), *Naive Bayes* dan *CART* [7]. Salah satu algoritma yang banyak diteliti adalah *Support Vector Machine* (SVM). SVM merupakan suatu teknik untuk menemukan *hyperplane* yang bisa memisahkan dua set data dari dua kelas yang berbeda [8]. SVM dikenal sebagai teknik *machine learning* yang dapat digunakan untuk menemukan fungsi pemisah

(*classifier*) yang optimal. Metode ini digunakan karena mampu melakukan klasifikasi data yang bisa dipisahkan oleh suatu garis linier maupun nonlinear. SVM merupakan salah satu metode terbaik yang dapat digunakan dalam masalah klasifikasi pola serta kemampuan generalisasi, yaitu kemampuan untuk mengklasifikasikan data lain yang tidak termasuk dalam data yang dipakai pada metode *machine learning* [9]. Selanjutnya akan dilakukan klasifikasi menggunakan *Naïve Bayes Classifier* (NBC) sebagai perbandingan terhadap metode SVM. NBC adalah salah satu klasifikasi probabilistik paling sederhana yang didasarkan pada teorema Bayes dan mencari peluang terbesar dari kemungkinan klasifikasi. NBC dapat melakukan proses klasifikasi dengan lebih cepat dibandingkan dengan metode lain yang lebih kompleks [10]. Meskipun termasuk dalam klasifikasi yang sederhana, namun dalam beberapa penelitian menunjukkan bahwa metode NBC memiliki performa yang baik dibandingkan metode lain yang lebih kompleks yaitu memiliki nilai akurasi lebih tinggi dibandingkan SVM dan *Random Forest* [11].

II. TINJAUAN PUSTAKA

A. Klasifikasi

Klasifikasi adalah proses penemuan model (fungsi) yang menggambarkan dan membedakan kelas data yang bertujuan agar bisa digunakan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [12]. Algoritma klasifikasi yang banyak digunakan secara luas, yaitu *decision/classification trees*, *bayesian classifiers/naive bayes classifiers*, *neural networks*, algoritma genetika, *rough sets*, *k-nearest neighbor*, metode *rule based*, *memory based reasoning*, dan *support vector machines* (SVM). Klasifikasi data terdiri dari 2 langkah proses. Pertama adalah *learning* (fase *training*), dimana algoritma klasifikasi dibuat untuk menganalisa data *training* lalu direpresentasikan dalam bentuk *rule* klasifikasi. Proses kedua adalah klasifikasi, dimana data *testing* digunakan untuk memperkirakan akurasi dari *rule* klasifikasi [12].

B. K-Fold Cross Validation

K-fold cross validation adalah suatu metode yang digunakan untuk mempartisi data menjadi data *training* dan data *testing*. sebuah teknik yang menggunakan keseluruhan *dataset* yang ada sebagai *training* dan *testing*. Metode ini banyak digunakan peneliti karena dapat mengurangi bias yang terjadi pada saat pengambilan sampel. *K-fold cross validation* secara berulang-ulang membagi data menjadi data *training* dan data *testing*, dimana setiap data mendapat kesempatan menjadi data *testing* [13]. Metode ini ditujukan agar diperoleh tingkat kepercayaan yang tinggi karena semua *dataset* dilibatkan sebagai data *training* maupun *testing* [14].

C. Fast Correlation Based Filter (FCBF)

Secara umum, suatu variabel dikatakan bagus jika variabel tersebut relevan dengan konsep kelas namun tidak redundan pada variabel yang lain. Pendekatan lain yang dilakukan yaitu memilih ukuran korelasi berdasarkan konsep *information theoretical entropy* [15]. *Entropy* dari variabel X didefinisikan pada persamaan (1).

$$H(X) = -\sum_{i=1}^n P(x_i) \log_2(P(x_i)) \quad (1)$$

Entropy variabel X jika diketahui variabel Y didefinisikan pada persamaan (2).

$$H(X|Y) = -\sum_{i=1}^n P(y_i) \sum_{j=1}^n P(x_j|y_i) \log_2(P(x_j|y_i)) \quad (2)$$

$P(x_i)$ adalah peluang posterior untuk semua nilai X dan $P(x_i|y_i)$ adalah peluang posterior dari X jika Y diketahui.

Berdasarkan *entropy* tersebut dapat diperoleh *Information Gain* pada persamaan (3).

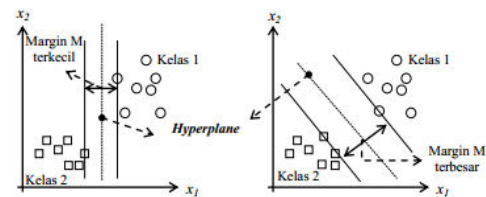
$$IG(X|Y) = H(X) - H(X|Y) \quad (3)$$

D. Support Vector Machine

Support Vector Machine (SVM) merupakan metode *machine learning* yang melakukan suatu teknik untuk menemukan fungsi pemisah (*classifier*) yang dapat memisahkan data menjadi dua kelas berbeda [8]. Tujuan dari SVM adalah untuk menemukan fungsi pemisah (*hyperplane*) terbaik diantara fungsi yang tidak terbatas jumlahnya untuk memisahkan dua buah kelas. SVM untuk klasifikasi dapat bekerja pada kasus klasifikasi linier maupun nonlinear.

1. Klasifikasi SVM Linier

Pada dasarnya, konsep SVM adalah linier *classifier* yaitu klasifikasi ke dalam 2 kelas yang secara linier dapat dipisahkan. Ilustrasi *hyperplane* pemisah dan margin SVM pada data yang dapat dipisahkan secara linier terdapat pada Gambar 1.



Gambar 1. *Hyperplane* SVM Linier.

Nilai margin merupakan nilai jarak terdekat *hyperplane* dengan data yang paling dekat dengan *hyperplane* tiap kelas. *Hyperplane* optimum diperoleh dengan memaksimalkan nilai margin. Secara matematis, formulasi permasalahan optimasi SVM untuk klasifikasi linier dalam primal space adalah:

$$L_p(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2} \mathbf{w}^T \mathbf{w} - \sum_{i=1}^M \alpha_i \left[y_i (\mathbf{w}^T \mathbf{x}_i + b) - 1 \right] \quad (4)$$

Nilai optimal pada persamaan (4) dapat dihitung dengan meminimalkan L_p terhadap \mathbf{w} dan b , serta memaksimalkan L_p terhadap α_i , sehingga pada persamaan (5) didapatkan formulasi optimasi SVM klasifikasi linier dalam dual space.

$$L_D(\mathbf{a}) = \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \quad (5)$$

Apabila data terpisah secara linier, maka fungsi pemisah atau *hyperplane* yang optimal adalah:

$$D(\mathbf{x}) = \sum \alpha_i y_i \mathbf{x}_i^T \mathbf{x} + b \quad (6)$$

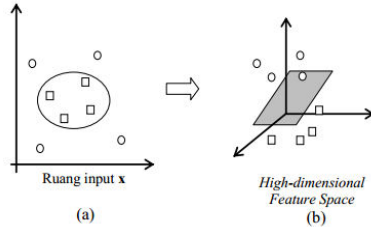
dengan nilai $b = \frac{1}{S} \sum (y_i - \mathbf{w}^T \mathbf{x}_i)$.

Selanjutnya data testing akan diklasifikasikan menjadi:

$$\mathbf{x} \in \begin{cases} \text{Kelas 1, jika } D(\mathbf{x}) > 0 \\ \text{Kelas 2, jika } D(\mathbf{x}) < 0 \end{cases} \quad (7)$$

2. Klasifikasi SVM Nonlinier

Pada umumnya masalah dalam dunia nyata jarang yang bersifat linier, akan tetapi kebanyakan bersifat nonlinier. Apabila data *training* tidak dapat dipisahkan secara linier, maka *classifier* yang diperoleh belum memiliki kemampuan generalisasi yang maksimal meskipun *hyperplane* yang diperoleh sudah optimal. Penyelesaian dalam masalah nonlinier tersebut, SVM dimodifikasi dengan memasukkan fungsi kernel. Ruang input dengan dua dimensi tidak dapat memisahkan data ke dalam dua kelas secara linier, sehingga dilakukan pemetaan vektor input oleh fungsi $\Phi(\mathbf{x})$ ke ruang vektor baru yang berdimensi lebih tinggi (3 dimensi).



Gambar 2. *Hyperplane* SVM Nonlinier.

Transformasi pada kasus SVM Nonlinier menggunakan fungsi kernel untuk memberikan kemudahan menentukan *support vector* pada SVM tidak perlu mengetahui bentuk dari pemetaan $\phi(\mathbf{x})$ yang sebenarnya [15]. Fungsi kernel yang umum digunakan pada metode SVM adalah:

1. Kernel Linier

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$$

2. Kernel Polynomial

$$K(\mathbf{x}_i, \mathbf{x}_j) = (-\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^p, \gamma > 0$$

3. Kernel Radial Basis Function (RBF)

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right)$$

4. Sigmoid Kernel

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(-\gamma \mathbf{x}_i^T \mathbf{x}_j + r)$$

Persamaan dual pada klasifikasi nonlinier diperoleh dari penerapan *kernel trick* pada persamaan (8).

$$L_D(\mathbf{a}) = \sum_{i=1}^M \alpha_i - \frac{1}{2} \sum_{i,j=1}^M \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (8)$$

Fungsi keputusan pada SVM nonlinier diperoleh melalui persamaan (2.16).

$$D(\mathbf{x}) = \sum_{i \in S} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}_j) + b \quad (9)$$

dengan $K(\mathbf{x}_i, \mathbf{x})$ adalah fungsi kernel. Selanjutnya data *testing* diklasifikasikan menggunakan fungsi keputusan:

$$\mathbf{x} \in \begin{cases} \text{Kelas 1, jika } D(\mathbf{x}) > 0 \\ \text{Kelas 2, jika } D(\mathbf{x}) < 0 \end{cases} \quad (10)$$

E. Naïve Bayes Classifier

Naïve Bayes merupakan perhitungan *teorema bayes* yang paling sederhana, karena mampu mengurangi kompleksitas komputasi dengan asumsi independensi yang kuat (naif) antar satu kelas dengan kelas yang lain. Secara umum, persamaan *Naïve Bayes* dapat dinotasikan pada persamaan (11) [16]:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (11)$$

Misalkan diberikan x_1, x_2, \dots, x_p adalah atribut yang digunakan untuk menentukan kelas y . Perhitungan *posterior probability* untuk setiap kelas y_i menggunakan teorema Bayes pada persamaan (12) [16].

$$P(y_i | x_1, \dots, x_p) = \frac{P(y_i)P(x_1, \dots, x_p | y_i)}{P(x_1, \dots, x_p)} \quad (12)$$

Persamaan 2.26 diperlukan perhitungan $P(x_1, \dots, x_p | y_i)$ dan setiap atribut diasumsikan saling bebas untuk kelas y , maka:

$$P(x_1, x_2, \dots, x_p) = P(x_1 | y)P(x_2 | y) \dots (x_p | y) \quad (13)$$

F. Confusion Matrix

Confusion matrix adalah alat yang digunakan untuk menganalisis seberapa baik *classifier* mengenali data dari kelas yang berbeda.

Tabel 1.

| Confusion matrix | | |
|------------------|---------------|---------------|
| Aktual | Prediksi | |
| | $\hat{\pi}_1$ | $\hat{\pi}_2$ |
| $\hat{\pi}_1$ | n_{11} | n_{12} |
| $\hat{\pi}_2$ | n_{21} | n_{22} |

Confusion matrix pada tabel 1 digunakan pada beberapa pengukuran performa klasifikasi yaitu akurasi, sensitivitas, spesifitas, dan presisi [12].

$$\text{akurasi} = \frac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22}} \quad (14)$$

$$\text{sensitivitas} = \frac{n_{11}}{n_{11} + n_{12}} \quad (15)$$

$$\text{spesifitas} = \frac{n_{22}}{n_{22} + n_{21}} \quad (16)$$

Akurasi klasifikasi data yang *unbalance* dapat dihitung menggunakan *Geometric-mean (G-mean)* dan *Area Under ROC Curve (AUC)* [17].

$$Gmean = \left(\prod_{i=1}^g R_i \right) = \sqrt{\text{sensitivitas} \times \text{spesifitas}} \quad (17)$$

$$AUC = \frac{1}{g} \sum_{i=1}^g R_i = \frac{1}{2} (\text{sensitivitas} + \text{spesifitas}) \quad (18)$$

G. Energi Listrik Sektor Rumah Tangga

PT Perusahaan Listrik Negara (Persero) merupakan perusahaan yang berusaha di bidang tenaga listrik untuk kepentingan masyarakat dan Negara [1]. Pelanggan PT PLN (Persero) dibagi menjadi 5 golongan berdasarkan keperluan pemakaian antara lain sektor rumah tangga, industri, bisnis, sosial, dan penerangan umum. Pelanggan rumah tangga dengan daya 900 VA dibagi menjadi pelanggan subsidi dan pelanggan rumah tangga mampu non subsidi.

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan pada penelitian ini merupakan data sekunder yang diperoleh dari PT. PLN (Persero) Distribusi Jawa Timur dan Badan Perencanaan Kota Surabaya tahun 2017. Data yang diperoleh dari PLN hanya berupa data pelanggan subsidi 900 VA, selanjutnya data ini dilakukan penyesuaian dengan data fakir miskin dari Bappeko dan diperoleh data pelanggan yang sesuai sebanyak 28949

pelanggan. Sedangkan data pelanggan rumah tangga subsidi 450 VA seluruhnya diambil langsung dari data fakir miskin Bappeko dan diperoleh sebanyak 50315 pelanggan. Total pelanggan rumah tangga yang akan digunakan pada penelitian ini berjumlah 79264 pelanggan.

B. Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian ini disajikan dalam Tabel 2. Variabel respon yang digunakan terdiri dari 2 kategori yaitu pelanggan rumah tangga subsidi listrik 450 VA dan pelanggan subsidi 900 VA dan terdapat 16 variabel prediktor.

Tabel 2.
Variabel Penelitian

| Variabel | Nama variabel | Label | Skala |
|-----------------|----------------------------------|--|---------|
| Y | Daya listrik | 1) subsidi 450 VA 2) subsidi 900 VA | Ordinal |
| X ₁ | Kepemilikan tempat tinggal | 1) bebas sewa/dinas 2) kontrak/sewa 3) milik sendiri | Ordinal |
| X ₂ | Luas bangunan | 1) $\leq 36 \text{ m}^2$ 2) $36 \text{ m}^2 < \text{luas} \leq 70 \text{ m}^2$ 3) $> 70 \text{ m}^2$ | Ordinal |
| X ₃ | Jenis lantai terluas | 1) tanah 2) semen/bata merah 3) ubin/tegel/teraso 4) keramik 5) marmer/granit | Ordinal |
| X ₄ | Jenis dinding terluas | 1) tembok 2) kayu 3) bambu | Ordinal |
| X ₅ | Jenis atap terluas | 1) seng 2) asbes 3) genteng tanah liat 4) genteng metal 5) genteng keramik 6) beton/genteng beton | Ordinal |
| X ₆ | Jumlah kamar tidur | 1) jumlah 0 – 1 kamar 2) jumlah 2 – 3 kamar 3) jumlah 4 – 5 kamar | Ordinal |
| X ₇ | Sumber air minum | 1) Sumur terlindung 2) Sumur bor/pompa 3) Leding eceran 4) Leding meteran 5) Air isi ulang 6) Air bermerk | Ordinal |
| X ₈ | Bahan bakar memasak | 1) Listrik 2) Gas >3kg 3) Gas 3kg 4) Gas kota/biogas 5) Minyak tanah 6) Kayu bakar | Ordinal |
| X ₉ | Kepemilikan lemari es | 1) Tidak 2) Ya | Ordinal |
| X ₁₀ | Kepemilikan televisi | 1) Tidak 2) Ya | Ordinal |
| X ₁₁ | Kepemilikan komputer atau laptop | 1) Tidak 2) Ya | Ordinal |
| X ₁₂ | Kepemilikan sepeda | 1) Tidak 2) Ya | Ordinal |
| X ₁₃ | Kepemilikan sepeda motor | 1) Tidak 2) Ya | Ordinal |
| X ₁₄ | Peserta program raskin | 1) Ya 2) Tidak | Ordinal |
| X ₁₅ | Jumlah anggota keluarga | 1) 1 – 2 orang 2) 3 – 4 orang 3) 5 – 6 orang 4) ≥ 7 orang | Ordinal |
| X ₁₆ | Status kesejahteraan | 1) 10% terendah 2) 11% - 20% terendah 3) 21% -30% terendah 4) 31% - 40% terendah | Ordinal |

C. Langkah Analisis

Langkah analisis yang digunakan pada penelitian ini adalah sebagai berikut:

1. Menganalisis karakteristik data pelanggan rumah tangga yang memperoleh subsidi listrik daya 450 VA dan pelanggan subsidi 900 VA
2. Melakukan seleksi variabel menggunakan metode FCBF, sehingga akan terbentuk 2 kelompok data yaitu data menggunakan seluruh variabel prediktor dan data hanya menggunakan variabel yang terseleksi. Selanjutnya 2 kelompok data tersebut masing-masing dibagi menjadi data *training* dan data *testing*
3. Pembagian data *testing* dan data *training* dengan metode *k-fold cross validation* dengan melakukan proses *folding* sebanyak 10 kali. Pada setiap pengujian, 1 *subset* digunakan sebagai data *testing* sebanyak 7927 data, sedangkan *subset* yang lain digunakan sebagai data *training* sebanyak $79264 - 7927 = 71337$ data
4. Melakukan klasifikasi dengan metode *Support Vector Machine* melalui 2 tahapan yaitu menggunakan semua variabel prediktor dan menggunakan seleksi variabel dengan metode FCBF. Fungsi kernel yang digunakan yaitu pada penelitian ini menggunakan fungsi kernel Linier dan *Radial Basis Function* (RBF). *Range* nilai parameter Cost dan Gamma pada penelitian ini menggunakan nilai $C = 0,1, 1, 10$, dan 100 dan nilai $\gamma = 0,1, 1, 10$, dan 100 [17].
5. Melakukan klasifikasi menggunakan *Naïve Bayes Classifier* melalui 2 tahapan yaitu menggunakan semua variabel prediktor dan menggunakan seleksi variabel dengan metode FCBF.
6. Membandingkan performansi antara metode *Support Vector Machine* dan *Naïve Bayes Classifier* berdasarkan tingkat akurasi ketepatan klasifikasi.

IV. ANALISIS DAN PEMBAHASAN

A. Karakteristik Pelanggan Subsidi Listrik

Karakteristik pembeda antara pelanggan rumah tangga subsidi listrik daya 450 VA dan 900 VA yaitu terdapat pada variabel jumlah kamar tidur, kepemilikan lemari es atau kulkas, kepemilikan sepeda, dan status kesejahteraan.

Tabel 4.
Jumlah Kamar Tidur

| Kamar | 450 VA | | 900 VA subsidi | |
|-------|--------|------|----------------|------|
| | Jumlah | % | Jumlah | % |
| 0 – 1 | 32262 | 64,1 | 11542 | 39,9 |
| 2 – 3 | 17521 | 34,8 | 16636 | 57,5 |
| > 3 | 532 | 1,1 | 771 | 2,7 |
| Total | 50315 | 100 | 28949 | 100 |

Berdasarkan Tabel 4, pelanggan subsidi daya 450 VA dengan persentase 64,1% memiliki jumlah kamar tidur hanya sebanyak 1 kamar atau tidak memiliki kamar tidur dalam tempat tinggalnya, sedangkan untuk kategori pelanggan 900 VA subsidi dengan persentase paling tinggi sebesar 57,5% memiliki jumlah kamar tidur sebanyak 2 atau 3 kamar.

Tabel 5.
Kepemilikan Lemari Es

| Kepemilikan | 450 VA | | 900 VA subsidi | |
|-------------|--------|------|----------------|------|
| | Jumlah | % | Jumlah | % |
| Tidak | 34183 | 67,9 | 13755 | 47,5 |
| Ya | 16132 | 32,1 | 15194 | 52,5 |

Berdasarkan Tabel 5, pelanggan 450 VA sebagian besar tidak memiliki lemari es yaitu dengan persentase 67,9%. Sedangkan untuk pelanggan daya 900 VA subsidi sebesar 52,5% dari total pelanggannya sudah memiliki lemari es, namun masih terdapat 47,5% total pelanggannya yang belum memiliki lemari es.

Tabel 6.

| Kepemilikan Sepeda | | | | |
|--------------------|--------|------|----------------|------|
| Kepemilikan | 450 VA | | 900 VA subsidi | |
| | Jumlah | % | Jumlah | % |
| Tidak | 31408 | 62,4 | 12643 | 43,7 |
| Ya | 18907 | 37,6 | 16306 | 56,3 |

Berdasarkan Tabel 6, sebagian besar pelanggan daya 450 VA dengan persentase 62,4% tidak memiliki sepeda, sedangkan pelanggan subsidi daya 900 VA sudah memiliki sepeda dengan persentase pelanggan 56,3%.

Tabel 7.

| Status Kesejahteraan | | | | |
|----------------------|--------|------|----------------|------|
| Status | 450 VA | | 900 VA subsidi | |
| | Jumlah | % | Jumlah | % |
| 10% terendah | 12249 | 24,3 | 3394 | 11,7 |
| 11% – 20% terendah | 14169 | 28,2 | 7100 | 24,5 |
| 21% – 30% terendah | 12918 | 25,7 | 8413 | 29,1 |
| 31% – 40 % terendah | 10979 | 21,8 | 10042 | 34,7 |

Berdasarkan Tabel 7, jumlah pelanggan 450 VA hampir merata pada masing-masing kategori dengan persentase paling tinggi sebesar 28,2% dengan status kesejahteraan 11% - 20% terendah, sedangkan untuk kategori pelanggan 900 VA dengan persentase paling tinggi sebesar 34,7% berada pada status kesejahteraan 31% - 40% terendah.

B. Seleksi Variabel Menggunakan FCBF

Hasil seleksi variabel menggunakan FCBF terdapat 7 variabel yang terseleksi antara lain jumlah kamar tidur, kepemilikan lemari es, status kepemilikan tempat tinggal, status kesejahteraan, luas lantai, kepemilikan sepeda, dan jenis lantai terluas. Variabel yang terseleksi tersebut juga akan digunakan pada tahapan klasifikasi dan akan dilakukan perbandingan dengan klasifikasi seluruh variabel prediktor.

C. Klasifikasi Support Vector Machine Menggunakan Seluruh Variabel Prediktor

Pada penelitian ini, metode SVM menggunakan fungsi kernel Linier dan kernel *Radial Basis Function* (RBF). *Range* nilai parameter C dan γ pada penelitian ini menggunakan nilai $C = 0,1, 1, 10$, dan 100 dan nilai $\gamma = 0,1, 1, 10$, dan 100 .

Pemilihan parameter optimal berdasarkan nilai rata-rata akurasi, AUC, dan G -Mean yang paling tinggi [21]. Hasil dari Tabel 8 dapat diketahui bahwa rata-rata nilai akurasi, AUC, dan G -Mean paling tinggi terdapat pada nilai $C = 0,1$ dan $C = 1$, selanjutnya dipilih nilai $C = 0,1$ untuk digunakan pada data *testing*. Dengan menerapkan model optimal pada data *testing*, maka akan diperoleh performa klasifikasi berdasarkan rata-rata dari nilai akurasi, AUC, dan G -Mean [18].

Hasil dari 10 percobaan yang dilakukan pada data *testing* pada Tabel 9, rata-rata nilai akurasi yang diperoleh dengan menggunakan kernel linier hanya sebesar 69%, dengan nilai AUC 63,5% dan nilai G -Mean sebesar 60%. Performa klasifikasi SVM kernel linier memberikan hasil yang kurang

baik untuk mengklasifikasikan pelanggan subsidi listrik.

Tabel 8.

Rata-rata Akurasi, AUC, dan G -Mean Kernel Linier (*Training*)

| Cost (C) | Akurasi (%) | AUC (%) | G -Mean (%) |
|--------------|-------------|---------|---------------|
| 0,01 | 69,0 | 63,2 | 59,5 |
| 0,1 | 69,0 | 63,3 | 59,6 |
| 1 | 69,0 | 63,3 | 59,6 |
| 10 | 69,0 | 63,2 | 59,5 |
| 100 | 68,9 | 63,1 | 59,5 |

Tabel 9.

Nilai Akurasi, AUC, dan G -Mean Kernel Linier (*Testing*)

| <i>Fold</i> | Akurasi (%) | AUC (%) | G -Mean (%) |
|-------------|-------------|---------|---------------|
| 1 | 68,7 | 63,5 | 60,3 |
| 2 | 69,4 | 63,6 | 60,2 |
| 3 | 69,2 | 64,0 | 60,8 |
| 4 | 69,9 | 64,3 | 61,3 |
| 5 | 68,7 | 63,3 | 59,8 |
| 6 | 69,4 | 63,9 | 60,5 |
| 7 | 68,2 | 62,8 | 59,0 |
| 8 | 68,6 | 62,8 | 59,2 |
| 9 | 69,2 | 63,4 | 59,6 |
| 10 | 68,9 | 63,4 | 59,9 |
| Rata-rata | 69,0 | 63,5 | 60,0 |

Tabel 10.

Rata-rata Akurasi, AUC, dan G -Mean Kernel RBF (*Training*)

| Cost (C) | Gamma (γ) | Akurasi (%) | AUC (%) | G -Mean (%) |
|--------------|--------------------|-------------|-------------|---------------|
| 0,1 | 0,1 | 72,4 | 66,2 | 62,2 |
| 0,1 | 1 | 76,3 | 68,8 | 62,8 |
| 0,1 | 10 | 75,7 | 67,6 | 60,7 |
| 0,1 | 100 | 75,7 | 67,6 | 60,7 |
| 1 | 0,1 | 76,8 | 72,3 | 70,3 |
| 1 | 1 | 86,8 | 84,8 | 84,5 |
| 1 | 10 | 87,3 | 85,8 | 85,6 |
| 1 | 100 | 87,3 | 85,8 | 85,6 |
| 10 | 0,1 | 81,4 | 78,2 | 77,3 |
| 10 | 1 | 87,3 | 85,9 | 85,7 |
| 10 | 10 | 87,3 | 85,8 | 85,6 |
| 10 | 100 | 87,3 | 85,8 | 85,6 |
| 100 | 0,1 | 84,8 | 80,2 | 78,9 |
| 100 | 1 | 87,3 | 85,9 | 85,7 |
| 100 | 10 | 87,3 | 85,8 | 85,6 |
| 100 | 100 | 87,3 | 85,8 | 85,6 |

Hasil performa klasifikasi kernel RBF pada Tabel 10 dapat diketahui bahwa untuk mendapatkan hasil klasifikasi yang baik nilai parameter C dan γ yang harus digunakan adalah $C = 10$ dan 100 pada $\gamma = 1$. Selanjutnya dipilih nilai $C = 10$ dan $\gamma = 1$ karena rata-rata nilai akurasi, AUC, dan G -Mean pada kombinasi tersebut telah memberikan hasil yang optimal dengan nilai tertinggi.

Tabel 11.

Nilai Akurasi, AUC, dan G -Mean Kernel RBF (*Testing*)

| <i>Fold</i> | Akurasi (%) | AUC (%) | G -Means (%) |
|-------------|-------------|---------|----------------|
| 1 | 87,0 | 85,6 | 85,4 |
| 2 | 87,6 | 86,3 | 86,1 |
| 3 | 87,0 | 85,7 | 85,6 |
| 4 | 88,1 | 86,8 | 86,6 |
| 5 | 87,9 | 86,6 | 86,4 |
| 6 | 86,9 | 85,4 | 85,2 |
| 7 | 87,7 | 86,4 | 86,2 |
| 8 | 87,3 | 86,0 | 85,9 |
| 9 | 87,3 | 86,0 | 85,9 |
| 10 | 79,5 | 77,2 | 76,7 |
| Rata-rata | 86,6 | 85,2 | 85,0 |

Hasil dari 10 percobaan yang dilakukan pada data

testing, rata-rata nilai akurasi yang diperoleh sebesar 86,6% dengan nilai AUC 85,2% dan nilai *G-Mean* sebesar 85,0%.. Performa klasifikasi berdasarkan nilai akurasi, AUC, dan *G-Mean* telah memberikan hasil yang cukup baik yaitu sudah mencapai diatas 80%.

D. Klasifikasi Support Vector Machine Hasil Seleksi FCBF

Dalam tahapan ini juga menggunakan fungsi kernel Linier dan kernel *Radial Basis Function* (RBF). Range nilai parameter *C* dan γ pada penelitian ini menggunakan nilai *C* = 0,1, 1, 10, dan 100 dan nilai γ = 0,1, 1, 10, dan 100.

Tabel 12.

| Hasil Kombinasi Parameter Kernel Linier (Training) | | | |
|--|-------------|-------------|-------------|
| Cost (C) | Akurasi (%) | AUC (%) | G-Mean (%) |
| 0,01 | 68,7 | 63,3 | 60,0 |
| 0,1 | 68,7 | 63,3 | 60,0 |
| 1 | 68,7 | 63,3 | 60,0 |
| 10 | 68,7 | 63,3 | 60,0 |
| 100 | 68,6 | 63,3 | 60,0 |

Hasil dari Tabel 12 dapat diketahui bahwa rata-rata nilai akurasi, AUC, dan *G-Mean* pada semua nilai *C* menunjukkan nilai yang sudah konvergen, selanjutnya dipilih nilai *C* = 0,1 untuk digunakan pada 10 data testing.

Tabel 13.

| Hasil Kombinasi Parameter Kernel Linier (Testing) | | | |
|---|---------------|---------|------------|
| Fold | Akurasi i (%) | AUC (%) | G-Mean (%) |
| 1 | 68,3 | 63,2 | 60,1 |
| 2 | 69,2 | 63,6 | 60,3 |
| 3 | 69,5 | 64,2 | 61,4 |
| 4 | 68,8 | 63,8 | 60,7 |
| 5 | 68,1 | 62,8 | 59,5 |
| 6 | 69,1 | 63,7 | 60,3 |
| 7 | 68,2 | 62,8 | 59,1 |
| 8 | 68,3 | 62,7 | 59,3 |
| 9 | 68,9 | 63,2 | 59,6 |
| 10 | 68,8 | 63,3 | 59,8 |
| Rata-rata | 68,7 | 63,3 | 60,0 |

Hasil dari 10 percobaan yang dilakukan pada data testing, rata-rata akurasi yang diperoleh menggunakan kernel linier hanya sebesar 68,7% dengan nilai AUC 63,3% dan nilai *G-Mean* sebesar 60%. Performa klasifikasi SVM kernel linier setelah dilakukan seleksi variabel menghasilkan nilai akurasi, AUC, dan *G-Mean* yang tidak lebih baik apabila dibandingkan menggunakan seluruh variabel prediktor yang dilibatkan dalam klasifikasi.

Tabel 14.

| Hasil Kombinasi Parameter Kernel RBF (Training) | | | | |
|---|--------------------|-------------|--------------|--------------|
| Cost (C) | Gamma (γ) | Akurasi (%) | AUC (%) | G-Mean (%) |
| 0.1 | 0.1 | 69,0 | 62,24 | 57,02 |
| 0.1 | 1 | 71,0 | 65,12 | 61,37 |
| 0.1 | 10 | 71,1 | 65,30 | 61,60 |
| 0.1 | 100 | 71,1 | 65,30 | 61,60 |
| 1 | 0.1 | 69,7 | 63,80 | 59,88 |
| 1 | 1 | 71,6 | 66,51 | 63,82 |
| 1 | 10 | 71,6 | 66,58 | 63,93 |
| 1 | 100 | 71,6 | 66,58 | 63,93 |
| 10 | 0.1 | 70,3 | 64,82 | 61,51 |
| 10 | 1 | 71,6 | 66,61 | 64,01 |
| 10 | 10 | 71,6 | 66,58 | 63,93 |
| 10 | 100 | 71,6 | 66,58 | 63,93 |
| 100 | 0.1 | 70,7 | 45,80 | 43,67 |
| 100 | 1 | 71,6 | 66,61 | 64,01 |

| | | | | |
|-----|-----|------|-------|-------|
| 100 | 10 | 71,6 | 66,58 | 63,93 |
| 100 | 100 | 71,6 | 66,58 | 63,93 |

Hasil Tabel 14 dengan kernel RBF dapat diketahui bahwa untuk mendapatkan hasil klasifikasi yang baik nilai parameter *C* dan γ yang harus digunakan adalah *C* = 10 dan *C* = 100 pada γ = 1. Selanjutnya dipilih nilai *C* = 10 dan γ = 1 karena rata-rata hasil akurasi, AUC, dan *G-Mean* pada kombinasi tersebut telah memberikan hasil yang optimal. Parameter optimal *C* = 10 dan γ = 1 diterapkan pada 10 fold data testing pada Tabel 15.

Hasil dari 10 percobaan yang dilakukan pada data testing, rata-rata nilai akurasi yang diperoleh sebesar 71,5% dengan nilai AUC 66,8% dan nilai *G-Mean* sebesar 64,4%. Performa klasifikasi SVM kernel RBF setelah dilakukan seleksi variabel menghasilkan nilai akurasi, AUC, dan *G-Mean* yang tidak lebih baik dibandingkan menggunakan seluruh variabel prediktor.

Tabel 15.

| Hasil Kombinasi Parameter Kernel RBF (Testing) | | | |
|--|-------------|---------|------------|
| Fold | Akurasi (%) | AUC (%) | G-Mean (%) |
| 1 | 69,9 | 65,4 | 63,1 |
| 2 | 72,4 | 67,2 | 64,6 |
| 3 | 71,6 | 67,3 | 65,2 |
| 4 | 72,2 | 67,5 | 65,5 |
| 5 | 71,4 | 66,7 | 64,4 |
| 6 | 71,3 | 66,4 | 63,9 |
| 7 | 71,2 | 66,4 | 63,7 |
| 8 | 71,3 | 66,4 | 63,8 |
| 9 | 71,9 | 67,0 | 64,5 |
| 10 | 72,0 | 67,4 | 65,0 |
| Rata-rata | 71,5 | 66,8 | 64,4 |

E. Klasifikasi Metode Naïve Bayes Classifier

Partisi data yang digunakan sama dengan metode SVM yaitu membagi data training dan testing menjadi 10 fold.

Tabel 16.

| Hasil Klasifikasi Naïve Bayes Classifier Seluruh Variabel | | | |
|---|-----------|------|--------|
| Fold | Akurasi i | AUC | G-Mean |
| 1 | 66,9 | 64,5 | 63,8 |
| 2 | 67,9 | 65,3 | 64,6 |
| 3 | 67,8 | 65,5 | 64,9 |
| 4 | 67,9 | 65,6 | 65,1 |
| 5 | 66,8 | 64,4 | 63,8 |
| 6 | 67,3 | 64,6 | 63,8 |
| 7 | 67,0 | 64,5 | 63,7 |
| 8 | 66,6 | 63,8 | 62,9 |
| 9 | 67,5 | 64,9 | 64,2 |
| 10 | 67,3 | 64,9 | 64,2 |
| Rata-rata | 67,3 | 64,8 | 64,1 |

Hasil dari Tabel 16 diketahui bahwa dari 10 percobaan yang dilakukan pada data testing, rata-rata nilai akurasi yang diperoleh adalah 67,3% dengan nilai AUC 64,8% dan nilai *G-Mean* sebesar 64,1%. Performa klasifikasi *Naïve Bayes Classifier* berdasarkan nilai akurasi, AUC, dan *G-Mean* memberikan hasil yang rendah pada kasus klasifikasi pelanggan rumah tangga subsidi listrik.

Tabel 17.

| Hasil Klasifikasi Naïve Bayes Classifier Seleksi Variabel FCBF | | | |
|--|-----------|------|--------|
| Fold | Akurasi i | AUC | G-Mean |
| 1 | 68,8 | 64,1 | 63,3 |
| 2 | 68,5 | 65,4 | 64,5 |
| 3 | 67,6 | 65,1 | 64,4 |
| 4 | 67,3 | 64,7 | 64,0 |

| | | | |
|-----------|------|------|------|
| 5 | 67,1 | 64,6 | 63,9 |
| 6 | 67,6 | 64,6 | 63,6 |
| 7 | 67,0 | 64,1 | 63,1 |
| 8 | 66,6 | 63,5 | 62,4 |
| 9 | 67,7 | 64,6 | 63,7 |
| 10 | 67,5 | 64,7 | 63,8 |
| Rata-rata | 67,6 | 64,5 | 63,7 |

Hasil dari Tabel 17 diketahui bahwa dari 10 percobaan yang dilakukan pada data *testing*, rata-rata nilai akurasi yang diperoleh adalah 67,6% dengan nilai AUC 64,5% dan nilai *G-Mean* sebesar 63,7%. Performa klasifikasi *Naïve Bayes Classifier* setelah dilakukan seleksi variabel memberikan hasil akurasi, AUC, dan *G-Mean* lebih rendah dibandingkan dengan menggunakan seluruh variabel prediktor.

F. Perbandingan Metode Klasifikasi

Langkah selanjutnya adalah membandingkan performa yang dihasilkan dari metode SVM dan NBC yaitu dengan membandingkan nilai akurasi, AUC, dan *G-Mean*.

Tabel 18.
Perbandingan Akurasi

| Metode | Akurasi (%) | AUC (%) | <i>G-Mean</i> (%) |
|--------------------|-------------|---------|-------------------|
| SVM Linier | 69,0 | 63,5 | 60,0 |
| SVM RBF | 86,6 | 85,2 | 85,0 |
| SVM Linier Seleksi | 68,7 | 63,3 | 60,0 |
| SVM RBF Seleksi | 71,5 | 66,8 | 64,4 |
| NBC | 67,3 | 64,8 | 64,1 |
| NBC seleksi | 67,6 | 64,5 | 63,7 |

Tabel 18 dapat dilihat bahwa dari hasil pengukuran akurasi, AUC, dan *G-Mean* untuk metode SVM kernel RBF lebih baik dibandingkan metode SVM kernel linier, serta hasil klasifikasi menggunakan seluruh variabel prediktor lebih baik dibandingkan menggunakan hasil seleksi variabel FCBF. Hasil klasifikasi menggunakan metode NBC juga diperoleh performa klasifikasi lebih tinggi menggunakan seluruh variabel prediktor dibandingkan hanya menggunakan prediktor hasil seleksi FCBF. Perbandingan antara metode SVM dan NBC diperoleh hasil bahwa metode SVM lebih baik dibandingkan metode NBC. Nilai akurasi, AUC, dan *G-Mean* untuk metode SVM kernel RBF tanpa seleksi variabel sudah mencapai persentase diatas 80% dibandingkan metode lainnya yang memberikan hasil performansi hanya dengan persentase 60%.

Pada Tabel 18 memberikan hasil bahwa metode SVM kernel RBF tanpa seleksi variabel memberikan performa yang paling baik. Berdasarkan nilai parameter optimal yang telah diperoleh pada Tabel 10 sebesar $C = 10$ dan $\gamma = 1$, fungsi *hyperplane* yang terbentuk untuk klasifikasi pelanggan rumah tangga subsidi 450 VA dan pelanggan subsidi 900 VA adalah:

$$D(x) = \sum_{i \in S} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b$$

dimana fungsi kernel yang digunakan adalah *Radial Basis Function* (RBF) dengan parameter $\gamma = 1$, dengan rumus:

$$K(\mathbf{x}_i, \mathbf{x}) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2\right) \\ = \exp\left(-\|\mathbf{x}_i - \mathbf{x}\|^2\right)$$

sehingga fungsi *hyperplane* yang diperoleh menjadi:

$$D(x) = \sum_{i \in S} \alpha_i y_i \exp\left(-\|\mathbf{x}_i - \mathbf{x}\|^2\right) + b$$

$$D(x) = \sum_{i=1}^N \alpha_i y_i \exp\left(-\|\mathbf{x}_i - \mathbf{x}\|^2\right) + 0,6973412 \cdot$$

Hasil klasifikasi yang diperoleh apabila nilai $D(x) < 0$ maka dikategorikan kedalam kategori pelanggan rumah tangga subsidi 450 VA, sedangkan apabila $D(x) \geq 0$ maka dikategorikan pelanggan rumah tangga subsidi 900 VA.

Tabel 19.

Hasil *Confusion Matrix*

| Kelas asli | Kelas prediksi | | | | Jumlah |
|------------|----------------|------|--------|------|--------|
| | 450 VA | % | 900 VA | % | |
| 450 VA | 4668 | 91,6 | 429 | 8,4 | 5097 |
| 900 VA | 512 | 18,1 | 2318 | 81,9 | 2830 |

Tabel 19 menunjukkan bahwa kategori 450 VA sebanyak 91,6% rumah tangga diklasifikasikan dengan benar dan sisanya 8,4% rumah tangga diprediksi masuk ke dalam kategori subsidi 900 VA. Pada kategori subsidi 900 VA sebanyak 81,9% rumah tangga diklasifikasikan dengan benar dan sisanya 18,1% rumah tangga diprediksi masuk ke dalam kategori 450 VA.

Selanjutnya akan dilakukan karakteristik pelanggan rumah tangga dari hasil klasifikasi menggunakan SVM kernel RBF.

Tabel 20.

Karakteristik Jumlah Kamar Tidur

| Kelas aktual | Kelas prediksi | % | | |
|--------------|----------------|-------|-------|------|
| | | 0 – 1 | 2 – 3 | >3 |
| 450 VA | 450 VA | 66,20 | 32,84 | 0,96 |
| 450 VA | 900 VA | 32,40 | 66,90 | 0,70 |
| 900 VA | 900 VA | 33,74 | 63,03 | 3,23 |
| 900 VA | 450 VA | 61,13 | 37,89 | 0,98 |

Hasil karakteristik berdasarkan Tabel 20, pelanggan rumah tangga subsidi 450 yang diprediksi menjadi pelanggan subsidi 900 VA sebesar 66,9% sudah memiliki jumlah kamar sebanyak 2 hingga 3 kamar. Hal ini sesuai dengan karakteristik yang dimiliki pelanggan subsidi 900 VA yang sebagian besar pelanggannya memiliki sebanyak 2 hingga 3 kamar. Pelanggan rumah tangga subsidi 900 yang diprediksi menjadi pelanggan subsidi 450 VA yaitu sebesar 61,13% yang sesuai dengan karakteristik yang dimiliki pelanggan subsidi 450 VA yaitu hanya memiliki 1 kamar atau tidak memiliki kamar tidur di dalam tempat tinggalnya.

Tabel 21.

Karakteristik Kepemilikan Lemari Es/Kulkas

| Kelas aktual | Kelas Prediksi | % | |
|--------------|----------------|--------------|--------------|
| | | Tidak | Ya |
| 450 VA | 450 VA | 70,72 | 29,28 |
| 450 VA | 900 VA | 43,36 | 56,64 |
| 900 VA | 900 VA | 41,59 | 58,41 |
| 900 VA | 450 VA | 64,65 | 35,35 |

Hasil karakteristik berdasarkan Tabel 21 pelanggan rumah tangga subsidi 450 VA yang diprediksi menjadi pelanggan subsidi 900 VA sebesar 56,643% sudah memiliki lemari es/kulkas. Hal ini sesuai dengan karakteristik yang dimiliki pelanggan subsidi 900 VA yang sebagian besar pelanggannya sudah memiliki lemari es/kulkas. Pelanggan rumah tangga subsidi 900 VA sebagian besar dianggap mendekati karakteristik pelanggan subsidi 450 VA yaitu 64,648% masih tidak memiliki lemari es/kulkas.

Tabel 22.

Karakteristik Kepemilikan Sepeda

| Kelas | Kelas | % |
|-------|-------|---|
|-------|-------|---|

| Aktual | Prediksi | Tidak | Ya |
|--------|----------|-------|-------|
| 450 VA | 450 VA | 64,31 | 35,69 |
| 450 VA | 900 VA | 46,62 | 53,38 |
| 900 VA | 900 VA | 38,18 | 61,82 |
| 900 VA | 450 VA | 63,09 | 36,91 |

Hasil karakteristik berdasarkan Tabel 22 pelanggan rumah tangga subsidi 450 yang salah diklasifikasikan menjadi pelanggan subsidi 900 VA sebanyak 53,38% memiliki sepeda. Hal ini sesuai dengan karakteristik yang dimiliki pelanggan subsidi 900 VA yang sebagian besar pelanggannya memiliki sepeda. Pelanggan rumah tangga subsidi daya 900 VA yang salah diklasifikasikan menjadi pelanggan subsidi 450 VA yaitu sebesar 63,086% pelanggan masih tidak memiliki sepeda.

Tabel 23.
Status Kesejahteraan

| Kelas aktual | Kelas prediksi | % | | | |
|-----------------|-------------------|----------|----------|----------|----------|
| | | Status 1 | Status 2 | Status 3 | Status 4 |
| 450 VA | 450 VA | 25,386 | 28,899 | 25,643 | 20,073 |
| 450 VA | 900 VA | 7,226 | 22,844 | 31,702 | 38,228 |
| 900 VA | 900 VA | 11,001 | 23,425 | 28,775 | 36,799 |
| 900 VA | 450 VA | 14,648 | 27,148 | 32,227 | 25,977 |

Hasil karakteristik berdasarkan Tabel 23 pelanggan rumah tangga subsidi 450 yang salah diklasifikasikan menjadi pelanggan subsidi 900 VA sebesar 38,228% memiliki status kesejahteraan 31% - 40% terendah, hal ini sesuai dengan karakteristik yang dimiliki pelanggan subsidi 900 VA. Pelanggan rumah tangga subsidi 900 VA yang salah di klasifikasikan menjadi pelanggan subsidi 450 VA sebesar 32,227% memiliki status kesejahteraan 21% - 30% terendah, hal ini mendekati karakteristik yang dimiliki pelanggan subsidi 450 VA sebagian besar dengan status kesejahteraan 11% - 20%.

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Hasil pengukuran akurasi, AUC, dan *G-mean* untuk metode SVM kernel RBF lebih baik dibandingkan metode SVM kernel linier, serta hasil klasifikasi menggunakan seluruh variabel prediktor lebih baik dibandingkan menggunakan hasil seleksi variabel FCBF. Hasil klasifikasi menggunakan metode NBC juga diperoleh performa klasifikasi lebih tinggi dengan menggunakan seluruh variabel prediktor dibandingkan menggunakan prediktor hasil seleksi FCBF. Perbandingan antara metode SVM dan NBC diperoleh hasil bahwa metode SVM lebih baik dibandingkan metode NBC dengan parameter optimal yang diperoleh adalah $C = 10$ dan $\gamma = 1$. Kategori 450 VA sebanyak 91,6% rumah tangga diklasifikasikan dengan benar dan sisanya 8,4% rumah tangga diprediksi masuk ke dalam kategori subsidi 900 VA. Pada kategori subsidi 900 VA sebanyak 81,9% rumah tangga diklasifikasikan dengan benar dan sisanya 18,1% rumah tangga diprediksi masuk ke dalam kategori 450 VA.

B. Saran

Saran yang dapat diberikan pada penelitian selanjutnya adalah lebih memperhatikan faktor atau variabel yang berpengaruh terhadap daya listrik rumah tangga yang memperoleh subsidi listrik, seperti faktor dari jam nyala, pemakaian kWh, tarif listrik, dan lain sebagainya. Metode yang digunakan bisa dikembangkan dengan metode-metode lain sebagai perbandingan terhadap metode *Support Vector Machine* dan *Naïve Bayes Classifier*. Perbandingan kernel pada klasifikasi SVM juga dapat digunakan selain kernel Linier dan RBF sebagai perbandingan untuk memperoleh hasil yang paling baik. Data yang digunakan untuk analisis klasifikasi lebih baik menggunakan data yang *balance* antar kategori kelasnya untuk meningkatkan performa klasifikasi.

DAFTAR PUSTAKA

- [1] S. Siregar and E. Warman, "Studi Prakiraan Kebutuhan Energi Listrik Tahun 2013-2017 Wilayah Kota Padang Sidimpuan dengan Metode Gabungan," *J. Singuda Ensikom*, vol. 1, no. 2, 2013.
- [2] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Creat. Inf. Technol. J.*, vol. 2, no. 3, 2015.
- [3] PLN, "Laporan Manajemen Perusahaan," Surabaya, 2017.
- [4] BPS Provinsi Jawa Timur, "Indikator Kesejahteraan Rakyat Jawa Timur 2017.," Surabaya, 2017.
- [5] Ditjen Ketenagalistrikan KESDM, "Statistik Ketenaga-listrikan 2017," Jakarta, 2016.
- [6] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed. San Fransisco: Morgan Kaufmann Publishers, 2011.
- [7] X. Wu and V. Kumar, "The Top Ten Algorithms in Data Mining," *J. Knowl. Inf. Syst.*, vol. 14, no. 1, pp. 1-37, 2009.
- [8] C. Cortes and V. Vapnik, "Support-Vector Networks," *J. Mach. Learn. Res.*, vol. 20, no. 3, pp. 273-297, 1995.
- [9] S. Gunn, "Support Vector Machines for Classification and Regression," Southampton, 1998.
- [10] H. Zhang, "The Optimality of Naïve Bayes," in *Proceeding of the Seventeenth International Florida Artificial Intelligence Research Society Conference*, 2004, pp. 562-567.
- [11] A. Nayak and D. Natarajan, "Comparative Study of Naïve Bayes Classifier in Sentiment Analysis of Twitter Feeds," *Int. J. Adv. Stud. Comput. Sci. Engineering*, vol. 5, no. 1, 2016.
- [12] J. Han, M. Kamber, and J. Pei, *Data Mining Concepts and Techniques*. Morgan Kaufman Publisher, 2012.
- [13] E. Gokgoz and A. Subasi, "Comparison of decision tree algorithms for EMG signal classification using DWT," *J. Biomed. Signal Process. Control*, vol. 18, pp. 138-144, 2015.
- [14] Y. Bengio, "No Unbiased Estimator of The Variance of K-Fold Cross Validation," *J. Mach. Learn. Res.*, vol. 5, pp. 1089-1105, 2004.
- [15] H. Liu and Y. Lei, "Feature Selection for High Dimensional Data: Fast Correlation Based Filter Solution," in *Proceeding of Twentieth International Conference on Machine Learning*, 2003, pp. 856-863.
- [16] S. Taheri and M. Mammadov, "Learning The Naïve Bayes Classifier with Optimization Models," *Int. J. Appl. Math. Comput. Sci.*, vol. 23, no. 4, 2013.
- [17] R. Sahitayakti and K. Fithriasari, "Klasifikasi Kesejahteraan Rumah Tangga di Provinsi Papua dengan Metode Regresi Logistik dan Support Vector Machine," *J. Sains dan Seni ITS*, vol. 4, no. 2, pp. 2337-3520, 2015.
- [18] C. Huang and C. J. Wang, "A GA-Based Feature Selection and Parameters Optimization for Support Vector Machines," *Int. J. Expert Syst. with Appl.*, vol. 31, no. 2, pp. 231-240, 2006.